# BIOGRAPHICAL INFORMATION

Andrew Zolnai
Petroleum/Pipeline Industry Manager
ESRI, Inc.

## Specific Responsibilities

Joined ESRI in 2000, is responsible for industry marketing in the petroleum and pipeline industry verticals.

## Past experience

1982-1986: Shell Canada and Geological Survey of Canada, exploration and field geologist
1986-1994: various, joint ventures and consulting, mapping and GIS in petroleum
1994-2000: Landmark Halliburton, support, training and project management in petroleum applications and GIS worldwide

## Educational Information

1980: B.Sc., University of Calgary, Calgary, Canada
1982: M.Sc., Queen's University, Kingston, Canada
Various continuing education in computing, GIS and project management

## Professional Memberships

AAPG (www.aapg.org), 1979–present, active practicing geologist
APEGGA (www.apegga.com), 1984-present, professional geologist

# STANDARDS AND METADATA
# - PART II -

Andrew Zolnai
ESRI, Inc.
380 New York Street, Redlands CA 92373
Gregory Gunther
USGS Energy

Designing a data model is the same basic process as designing any GIS database. Current activities include pipeline, petroleum, geological survey and land cadastre data models. Next comes codifying data about the data, its content, quality and condition: these are metadata that help find information, allow informed access and usage of data, and organize data logically. 70% of GIS cost is data, and must be efficiently located and evaluated from anywhere at any time. Producers must publish metadata, and users discover resources using metadata. A catalog brings producers and users together via metadata. Guidelines for this include: store and manage metadata as part of the dataset, support FGDC and ISO metadata standards, store metadata as XML, create metadata style sheets, and synchronize metadata. Metadata is an important and integral part of a creating a distributed GIS. The example metadata portal cited is the USGS Energy Program World Assessment. Metadata standards, data categorization and search page design were integral to their deployment. Key to future success will be to plan a catalog service, develop a vision, and identify goals. Metadata fuels the catalog service: make it part of your work flow, oversee its documentation, and arrange to document legacy data.

## DATA MODEL RECAP

This presentation is the second part of last year's, which outlined data models and their use in the pipeline industry. The following will be reviewed: data models, current activities, metadata, current spatial data infrastructure and future directions. Metadata help to extend the availability of data and the use of data models, and thus the successful creation of portals. This is key to the enterprise deployment of GIS.

Designing a Data Model is the same basic process as designing any GIS database: The conceptual design includes three iterative steps: first the collection of information (maps, applications, data sources, and metadata), and the identification of thematic layers (map scale, relationships, methods, and properties). Second the definition of each layer (feature representation, attributes, and symbology), and of the database structure (feature classes, relationships, domains, and rules). The physical design is to pilot, refine and implement each of the above iteratively, until one reaches a working, essential data model. The goal is the 20% of the data sets that achieve 80% of the work, rather than an exhaustive data standard that covers every detail (Figure 1).

Current Activities

These include the ArcGIS Pipeline Data Model (APDM, released last year and updated this year), the Public Petroleum Data Model (PPDM, work in progress), the FIG Cadastre Data Model (planned for 2014) and the Geological Data Model (also work in progress). These are essential data models that seek the 80/20 rule above. In fact, both APDM and PPDM have respectively Core and Light spatial implementations: these are mapped to the full data standards proposed by PODS (www.pods.org) and PPDM (www.ppdm.org) respectively (note that both associations arrived at the same process independently). A corollary to essential data models, is that they can be combined as business requires: rather than be s full data standard, pieces are linked according to usage (pipeline and oceanographic offshore, land and petroleum onshore, for example).

## METADATA

What it is

According to ISO (International Standards Organization), metadata is ‘data about data’. FGDC (US Federal Geographic Data Committee) proposes ‘data about content, quality, condition, and other characteristics of data’ (they also support ‘profiles’ that help in creating and transferring metadata). NASA-GCMD (Ground Control Meta Data), ‘descriptive information that characterizes a set of quantitative and/or qualitative measurements’. In this context, standards are content models that describe the "shape" of the metadata, not the format itself - in other words, this is to help describe and then find data according to predefined criteria, rather than be a full description of the data itself. Metadata thus allows to describe digital assets with consistency, through the use of common terminology, readable by people and by machines both.

Metadata essentially help find information - think of what populates a card catalog in a traditional library. It allows informed access and data usage across corporate systems, especially if they are diverse in content and dispersed in geography – they allow for example to store and retrieve legal statements regarding resources of data, in order to protect liability in data and other investments (this is for the recent US Sarbanes-Oxley Act vis-à-vis federal regulations, and the on-going Freedom of Information Act vis-à-vis the public). It helps organizes data logically, through the tools available to manage metadata. It enables resource decision management through the corporate discipline it helps impose in the data infrastructure (that is the management of how to allocate resources in information storage, access and dissemination).

What it does

Metadata thus save time and money. 70% of GIS costs revolve around data access. Data must be efficiently located and evaluated from anywhere at any time, and that means: Producers publish metadata, users discover resources using metadata, and the catalog: is the meeting place for producers and users via said metadata. And this is best done via distributed GIS. Here are a few guidelines we discovered in deploying metadata portals:

- Store and manage metadata as part of the dataset
- Keep data and metadata together - metadata travels with the data in the ESRI's geodatabase (spatially enabled database), and thus cannot get lost – if for example one export a resource, then the metadata accompanies the data
- One must preview metadata and the data together in the GIS in order to help manage it - this can be done via separate halves of the internet browser screen in http://geodata.gov (US government data portal for public access to its geographic data)
- Support FGDC and ISO metadata standards out-of-the-box, straight in the applications – this is the basis for COTS (commercial off the shelf) applications, which are far easier to maintain and to share via IT standards
- Include a metadata editor that is fully compliant with FGDC – this will enable others' access to the same metadata - also support the ISO Draft standard to add custom metadata pages
- Customize the metadata editors to access other standards – this is in anticipation of a need ot access other data that don't follow all standards, but tat appears ot have been quite common.

How it does it

Metadata are better stored as XML: It is an open industry standard. Its hierarchical structure is appropriate to create, store and retrieve metadata. It is oriented toward publishing and distribution - in particular over the Internet – and is machine-readable. Display definitions are for example stored separately in style sheets – it is thus more flexible when handling additional elements (items are always added in na iterative process) - metadata style sheets are displayed by applying an XML transformation (XSLT).

If data are stored in various and separate instances, it is important to synchronize the metadata across said instances – synchronizers will populate data set properties into specific metadata elements, such as bounding box and title. They are currently available in ISO, FGDC, and the Geography Network. One must then control which synchronizers are on and how they perform. Depending on what metadata standard is chosen, it helps to determine the need for metadata elements from both? Starting simple, and then keeping it so, are always recommended, as long as they achieve set goals.

Example

An example metadata service links ArcCatalog (ESRI's tool to view and manage, but not display and process, GIS data) to ArcIMS (ESRI Internet Map Server) that access data from ArcSDE (spatial data engine, that indexes data geographically in Oracle, DB2, SQL

Server etc.), and a GIS portal toolkit (ESRI tools that allow users to create portals, secure access points to corporate data over the internet or the intranet).

In summary, metadata is an important and integral part of a creating a distributed GIS. When setting up a Metadata Service, control who accesses and publishes it, when indexing will occur, and who will validate it when publishing. Note in closing that Metadata Services support Metadata Catalogs (internal to a corporation on its network) and Metadata Portals (the same on intranet or internet).

## METADATA PORTAL - USGS ENERGY PROGRAM WORLD ASSESSMENT

- Formal content type development and metadata requirements for each individual content type (which maps nicely to G.net)
- Full compliance with geospatial One-Stop and the Geography Network
- Spatial relevance scoring (work originally done by Ken Lanfear of USGS)
- Discovery by multiple means (that is, using many ways to access data)
- Over 1000 products that are indexed via individual metadata documents

<u>Metadata Standards – USGS Energy:</u>

- Based On FGDC Metadata Standard
- Formally defined Metadata Requirements for each Content Type
- Relationships defined in metadata using Larger Work Citation
    - ArcIMS Metadata Relationship then Programmatically inferred during publishing

<u>Data Categories – USGS Energy:</u>

- Future Extension of topical Categories based on Energy Industry Thesauri
- Extended g.Net Content Types to adequately categorize products within EP
- Extended Content Type Categories Include
    - *Online Reports-Chapters*
    - *Offline Reports-Chapters*
    - *Abstracts*
    - *Presentations*
    - *Digital Maps*
    - *Applications*
    - *Downloadable GIS Data*
    - *Tabular Data*
    - *Graphical Data*
    - *Live GIS Data and Maps*
    - *Geoprocessing Services*
    - *Documents*

# FUTURE DIRECTIONS

## Lessons learned

The question around standards is: to wait or not to wait? Data model templates are works-in-progress, but standards are needed to adhere to. So which one does one choose and on which criteria? Once that is done, users find data so fast that it's asking to "drinking from a fire-hose" – manage expectations and prepare robust workflows, because operationally, it is alive. So communicate, communicate… and find out where do users want to go today? If the metadata portal I.S.P. properly planned and implemented, then managers, developers and users will find out that it's not a UFO, it's an application.…

## Data Model to Metadata…

A few items will be suggested in linking metadata to data models, but these have not been acted upon yet by any organization. After data models stabilize (APDM, PPDM for example), they can be expanded upon in the domain of portals. Links developed among them (for ex. white papers among APDM and PODS), will help make them accessible to meta data catalog protocols (in other words, it's easier to create metadata on a data structure that is known and stable). Creating XML extensions, possibly working with POSC organisation (they already have connections to PPDM), will likely be a good place to start. Once they publish extensions (managing body yet to be determined), users will be able to create portals much more easily against their own essential data models already in place. Implementation of metadata may be boosted by one organisation sharing use case scenario, among key users who look to be champions. It may help to glean more best practices from other fields who already did this, such as document management.

## Metadata in the Enterprise

Metadata fuels the catalog service: Users will search a Metadata Service using ArcCatalog or Metadata Explorer. It will be easy to add data directly to a map, and the publish the attendant metadata in the catalog. Standard connectors (following, say, the Z39.50 protocol) will let people search Metadata Services as part of their own SDI Clearinghouse. GIS Catalog Portals will thus have an implementation infrastructure the user community can share.

# CONCLUSION:

The increased availability of data renews the importance of metadata standards and procedures. Quality metadata increases the confidence users have in the information and analysis they conduct. Metadata promotes the reuse and understanding of data over a longer period of time.

Figure 1:

Template-based GIS projects combine essential data models to meet specific needs, and avoid the use of very large, all-encompassing models.